

Asymmetrically extremely dilute neural networks with non-trivial dynamics

J P L Hatchett and A C C Coolen

Department of Mathematics, King's College, University of London, The Strand,
London WC2R 2LS, United Kingdom

Abstract. We study graded response attractor neural networks with asymmetrically extremely dilute interactions and Langevin dynamics. We solve our model in the thermodynamic limit using generating functional analysis, and find (in contrast to the binary neurons case) that even in statics one cannot eliminate the non-persistent order parameters. The macroscopic dynamics is driven by the (non-trivial) joint distribution of neurons and fields, rather than just the (Gaussian) field distribution. We calculate phase transition lines and present simulation results in support of our theory.

PACS numbers: 75.10.Nr, 05.20.-y, 64.60.Cn

E-mail: hatchett@mth.kcl.ac.uk, tcoolen@mth.kcl.ac.uk

1. Introduction

Hopfield-type models [1] are recurrent networks of binary neurons (or Ising spins) with specific types of pair-interactions, designed to store and retrieve information in a distributed way. The main interest in their properties relates to the regime of operation close to saturation, where the number of patterns stored p scales with the number of bonds per neuron c (i.e. $\alpha = p/c$ remains finite as $N \rightarrow \infty$). Over the years they have been studied intensively using statistical mechanical tools, in various connectivity versions. Rigorous analysis, based on techniques borrowed from spin-glass theory, has been carried out for Ising spin neural networks with full connectivity [3, 4, 5, 6] ($c, N \rightarrow \infty, c/N = 1$), and for asymmetric [7, 8] and symmetric extreme dilution [9] ($c, N \rightarrow \infty, c/N \rightarrow 0$). The approach of [7, 8] also proved effective for solving models with continuous neurons and discrete time dynamics [10]. More recently, attention has turned to finite connectivity attractor neural networks [11] ($c = \mathcal{O}(1), N \rightarrow \infty$).

In this paper we study the dynamics of asymmetrically extremely dilute graded response attractor neural networks with Langevin dynamics, near saturation. Such systems violate detailed balance in two ways, firstly due to the interaction asymmetry and secondly by the presence of a non-linear gain function (required for firing rates to saturate). Equilibrium statistical mechanics no longer applies; the system will never be in equilibrium. Yet, one still expects evolution to a stationary state, albeit one with microscopic probability currents. The stationary state of fully connected networks of

graded response neurons with continuous time dynamics [2] close to saturation have been studied at $T = 0$, [12, 13, 14] together with dynamics at arbitrary T but away from saturation [15]. In the present case of asymmetric extreme dilution one might expect the ideas of [7, 8] to apply, and anticipate the familiar simplification resulting from Gaussian distributed fields. This turns out to be wrong. In [7, 8] and [10] the spins were (stochastically) aligned to local fields in parallel, at discrete intervals. Spins at time $t + 1$ did not depend on spins at time t other than via the fields. Due to the asymmetric dilution, the field distribution was Gaussian on finite time-scales (characterized by two scalars), leading directly to simple iterative laws for one or two scalar order parameters. In contrast, although here the fields still have a Gaussian distribution, the Langevin dynamics makes our model qualitatively different: spin updates now depend strongly on present values, in addition to the fields, and we will therefore need the (non-Gaussian) joint spin-field distribution.

In this paper we use the generating functional analysis method of [16], which has a strong record in the area of asymmetric disordered spin systems (e.g. [17, 18, 19, 20]), following closely the exposition in [15] which allows us to be compact. As always, in the infinite size limit one finds an effective single spin equation, which in the special case of complete asymmetry will have no retarded self-interaction. This induces simplifications, resulting here in explicit self-consistent laws for order parameter functions. The further ansatz of time translation invariance leads to a tractable stationary state problem, from which, however, non-persistent order parameters cannot be fully removed. In addition to solving the full (exact) order parameter equations numerically, we also investigate an interpolative approximation that does depend on only a few scalar quantities and compare the predictions of both the full and the approximate theory with numerical simulation data. The interpolative theory is exact for zero α and also at zero T for small α (where small depends on the choice of gain function). We investigate the phase diagram of our model in the (α, T) plane, and examine the transition from non-recall to recall (we show that there is no spin-glass phase). Numerical simulations support our theoretical findings and predictions.

2. Model definitions

Our model describes a network of N continuous neurons $u_i \in \mathbb{R}$ which evolve in time according the following Langevin equation

$$\frac{d}{dt}u_i(t) = \sum_{j \neq i} J_{ij}g[u_j(t)] - u_i(t) + \eta_i(t), \quad (1)$$

where $g[u]$ is an odd sigmoidal function, which saturates for $u \rightarrow \pm\infty$ to ± 1 (respectively), and with the standard Gaussian white noise terms $\eta_i(t)$ with moments

$$\langle \eta_i(t) \rangle = 0, \quad \langle \eta_i(t)\eta_j(t') \rangle = 2T\delta_{ij}\delta(t - t'). \quad (2)$$

The non-negative parameter T controls the level of noise, with $T = 0$ and $T = \infty$ corresponding to deterministic and fully random dynamics, respectively. The process

(1) does not obey detailed balance (not even for symmetric J_{ij}), and thus will never reach equilibrium. This rules out the techniques of equilibrium statistical mechanics, and we are forced to solve the dynamics. In the fashion of attractor neural networks we store p (randomly drawn) patterns $\boldsymbol{\xi}^\mu = (\xi_1^\mu, \dots, \xi_N^\mu) \in \{-1, 1\}^N$ in this system, with $\mu = 1, \dots, p$, by choosing diluted Hebbian-type interactions $\{J_{ij}\}$ as in [7]:

$$J_{ij} = \frac{c_{ij}}{c} \sum_{\mu=1}^p \xi_i^\mu \xi_j^\mu \quad c_{ij} \in \{0, 1\} \quad (3)$$

The (fixed) $\{c_{ij}\}$ define the connectivity, and are drawn at random according to

$$i < j : \quad P(c_{ij}) = \frac{c}{N} \delta_{c_{ij},1} + (1 - \frac{c}{N}) \delta_{c_{ij},0} \quad (4)$$

$$i > j : \quad P(c_{ij}) = r \delta_{c_{ij},c_{ji}} + (1 - r) \left[\frac{c}{N} \delta_{c_{ij},1} + (1 - \frac{c}{N}) \delta_{c_{ij},0} \right] \quad (5)$$

The parameter c is seen to specify the average number of connections per spin, and r controls the symmetry of our architecture. We will choose extreme dilution, i.e. $\lim_{N \rightarrow \infty} c/N = \lim_{N \rightarrow \infty} c^{-1} = 0$, such as $c = \log(N)$. The number of patterns p to be stored is expected to scale with c , so we define $p = \alpha c$ with $\alpha > 0$ finite.

For simplicity, and in line with previous analyses, we will make the so-called ‘condensed ansatz’ in solving the dynamics: we assume that the system state has an $O(N^0)$ overlap only with a single pattern, say $\mu = 1$. This situation is induced by initial conditions: we take a randomly drawn $\mathbf{u}(0)$, generated by

$$p(\mathbf{u}(0)) = \prod_i \left\{ \frac{1}{2} [1 + m_0] \delta(u_i(0) - \xi_i^1) + \frac{1}{2} [1 - m_0] \delta(u_i(0) + \xi_i^1) \right\}, \quad (6)$$

with m_0 measuring the overlap between the initial state and the first pattern, $m_0 = N^{-1} \sum_i \langle \xi_i^1 u_i(0) \rangle$. The patterns $\mu > 1$ and the architecture variables $\{c_{ij}\}$, are viewed as disorder. We assume that for $N \rightarrow \infty$ the macroscopic behaviour of the system is ‘self-averaging’, i.e. only depends on the statistical properties of the disorder rather than its microscopic realization. Averages over the disorder will be written as $\overline{\dots}$.

3. Generating functional analysis

To analyze the dynamics following [16] we discretize time (in units of Δ), write the probability density of observing a microscopic ‘path’ $\{\mathbf{u}(0), \dots, \mathbf{u}(t_m)\}$ through phase space as $P[\mathbf{u}(0), \dots, \mathbf{u}(t_m)]$, and introduce the familiar generating function $Z_\Delta[\boldsymbol{\psi}]$:

$$\begin{aligned} Z_\Delta[\boldsymbol{\psi}] &= \overline{\langle \exp\{-i \sum_{i \leq N} \sum_{t \leq t_m} \Delta \psi_i(t) g[u_i(t)]\} \rangle} \\ &= \int d\mathbf{u}(0) \dots d\mathbf{u}(t_m) P[\mathbf{u}(0), \dots, \mathbf{u}(t_m)] \exp\{-i \sum_i \sum_t \Delta \psi_i(t) g[u_i(t)]\} \end{aligned} \quad (7)$$

Later we will put $\Delta \rightarrow 0^\ddagger$. Note that (7) differs from the standard definition in the appearance of the non-linear function $g[u]$ in the exponent. For full details on the

\ddagger The non-trivial technicalities related to the assumed commutation of the limits $N \rightarrow \infty$ and $\Delta \rightarrow 0$ are familiar issues in applying saddle-point arguments to path integrals, which we will not discuss here.

procedure for deriving from (7), in the limit $N \rightarrow \infty$, an effective single spin equation we refer to e.g. [15]. The discretized version of (1) describes a Markov process, so the path probability density is a product of individual transition probability densities. In the Itô convention we do not pick up a Jacobian term from the discretization. We add time-dependent external fields $\theta_i(t)\xi_i^1$ to the deterministic forces in order to define response functions later, and proceed in the standard manner to a saddle point problem. Solution of the latter leads us to a closed macroscopic theory in terms of the familiar dynamic order parameters $m(t) = \lim_{N \rightarrow \infty} N^{-1} \sum_i \xi_i^1 \overline{g[u_i(t)]}$ (the recall overlap), $C(t, t') = \lim_{N \rightarrow \infty} N^{-1} \sum_i \overline{g[u_i(t)]g[u_i(t')]}$ (the single-site correlation function) and $G(t, t') = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_i \partial g[u_i(t)] / \partial \xi_i^1 \theta_i(t')$ (the single-site response function). The order parameters are to be solved from the saddle-point equations

$$m(t) = \langle g[u(t)] \rangle_\star \quad (8)$$

$$C(t, t') = \langle g[u(t)]g[u(t')] \rangle_\star \quad (9)$$

$$G(t, t') = \partial \langle g[u(t)] \rangle_\star / \partial \theta(t') \quad (10)$$

Here the measure $\langle \dots \rangle_\star$ is defined by the statistics of the following effective single spin process (having set $\psi = 0$ and taken the limit $\Delta \rightarrow 0$):

$$\frac{d}{dt}u(t) = -u(t) + m(t) + \theta(t) + \alpha r \int_{-\infty}^t dt' G(t, t')g[u(t')] + \phi(t) \quad (11)$$

where $\phi(t)$ is a zero mean Gaussian process with covariance

$$\langle \phi(t)\phi(t') \rangle = 2T\delta(t - t') + \alpha C(t, t'). \quad (12)$$

In the remainder of this paper we will consider $r = 0$ only (i.e. fully asymmetric dilution, with c_{ij} independent of c_{ji}), so that the retarded self-interaction in (11) vanishes. This also implies that there is no longer a need to solve for the response function, as the macroscopic laws (8,9) now close already for the order parameters $\{m(t), C(t, t')\}$.

4. Solving the single spin equation for asymmetric dilution

As our main interest will be in stationary state properties, we choose initial conditions at $t_0 = -\infty$. For $r = 0$ we can readily integrate the stochastic equation (11), giving

$$u(t) = k(t) + Z(t) \quad (13)$$

$$k(t) = \int_{-\infty}^t dt' e^{-(t-t')} [m(t') + \theta(t')] \quad Z(t) = \int_{-\infty}^t dt' e^{-(t-t')} \phi(t') \quad (14)$$

So $Z(t)$ is also a zero mean Gaussian noise, now with covariance $\langle Z(t)Z(t') \rangle = \Xi(t, t')$

$$\Xi(t, t') = T e^{-|t-t'|} + \alpha \int_{-\infty}^t ds e^{-(t-s)} \int_{-\infty}^{t'} ds' e^{-(t'-s')} C(s, s') \quad (15)$$

We may use (13,14) and the standard abbreviation $Dz = (2\pi)^{-1/2} e^{-\frac{1}{2}z^2} dz$ for Gaussian measures to write the closed equations (8,9) in the following explicit form:

$$m(t) = \int Dx g[k(t) + x\sqrt{\Xi(t, t)}] \quad (16)$$

$$C(t, t') = \int Dx Dy g[k(t) + a_1(t, t')(a_2(t, t')x + a_3(t, t')y)] \\ \times g[k(t) + a_1(t, t')(a_2(t, t')x + a_4(t, t')y)] \quad (17)$$

in which (using the symmetry of Ξ):

$$a_1(t, t') = 1/\sqrt{\Xi(t, t) + \Xi(t', t') - 2\Xi(t, t')} \\ a_2(t, t') = \sqrt{\Xi(t, t)\Xi(t', t') - \Xi(t, t')^2} \\ a_3(t, t') = \Xi(t, t) - \Xi(t, t') \\ a_4(t, t') = \Xi(t', t) - \Xi(t', t')$$

Let us turn to time translation invariant solutions (TTI), where $m(t) = m$, $\theta(t) = \theta$ and $C(t, t') = C(t - t')$, and consequently $\Xi(t, t') = \Xi(t - t')$. This will simplify our analysis, and experience with asymmetric systems suggests that most, if not all, solutions will asymptotically obey TTI. We now find that equations (16,17) simplify to

$$m = \int Dx g[m + \theta + x\sqrt{\Xi(0)}] \quad (18)$$

$$C(\tau) = \int Dx Dy g \left[m + \theta + x\sqrt{\frac{1}{2}[\Xi(0) + \Xi(\tau)]} + y\sqrt{\frac{1}{2}[\Xi(0) - \Xi(\tau)]} \right] \\ \times g \left[m + \theta + x\sqrt{\frac{1}{2}[\Xi(0) + \Xi(\tau)]} - y\sqrt{\frac{1}{2}[\Xi(0) - \Xi(\tau)]} \right] \quad (19)$$

$$\Xi(\tau) = Te^{-|\tau|} + \frac{1}{2}\alpha \int_{-\infty}^{\infty} du e^{-|u+\tau|} C(u) \quad (20)$$

We note that, upon repeated differentiation with respect to τ , one can also derive from this latter equation the following identity:

$$\Xi''(\tau) = \Xi(\tau) - \alpha C(\tau) \quad (21)$$

In detailed balance models at this stage one would be able to transform away the non-persistent order parameters, and derive closed equations for the persistent objects m , $q_0 \equiv C(0)$ and $q \equiv C(\infty)$. Here such a reduction is not possible. To appreciate this we first rewrite (19) (by suitably transformations of integration variables) as

$$C(\tau) = \int Dx \left\{ \int Dy g \left[m + \theta + x\sqrt{\Xi(\tau)} + y\sqrt{\Xi(0) - \Xi(\tau)} \right] \right\}^2 \quad (22)$$

Using the relation $\Xi(\infty) = \alpha q$ and the short-hand $\Xi(0) = \kappa$, our expressions for the persistent quantities $\{m, q_0, q\}$ then take the form

$$m = \int Dx g[m + \theta + x\sqrt{\kappa}], \quad q_0 = \int Dx g^2[m + \theta + x\sqrt{\kappa}] \quad (23)$$

$$q = \int Dx \left\{ \int Dy g \left[m + \theta + x\sqrt{\alpha q} + y\sqrt{\kappa - \alpha q} \right] \right\}^2 \quad (24)$$

with

$$\kappa = T + \alpha \int_0^{\infty} d\tau e^{-\tau} C(\tau) \quad (25)$$

Unless $T = 0$ (where one has solutions with $q = q_0$ and $C(\tau) = q$ for all τ , to which we will return later), the term κ can not be expressed in terms of $\{m, q_0, q\}$, but requires knowledge of non-persistent correlations $C(\tau)$. From this stage onwards we will take $\theta = 0$, in order to study different solution types and transitions between them.

5. Special solutions

5.1. Non-recall TTI solutions

Non-recall solutions are those where $m = 0$ (to be expected for large T), and where in the TTI regime we are left only with a closed set of equations for $C(\tau)$:

$$C(\tau) = \int Dx Dy g \left[x \sqrt{\frac{1}{2}[\Xi(0) + \Xi(\tau)]} + y \sqrt{\frac{1}{2}[\Xi(0) - \Xi(\tau)]} \right] \\ \times g \left[x \sqrt{\frac{1}{2}[\Xi(0) + \Xi(\tau)]} - y \sqrt{\frac{1}{2}[\Xi(0) - \Xi(\tau)]} \right] \quad (26)$$

$$\Xi(\tau) = T e^{-|\tau|} + \frac{1}{2} \alpha \int_{-\infty}^{\infty} du e^{-|u+\tau|} C(u) \quad (27)$$

The simplest case is $T \rightarrow \infty$. Since $C(\tau)$ is bounded, we may expand our equations and find that the second term in (27) is negligible relative to the first. This results in

$$\lim_{T \rightarrow \infty} C(\tau) = \frac{2}{\pi} \arcsin(e^{-|\tau|}) \quad (28)$$

Clearly $\lim_{T \rightarrow \infty} C(0) = 1$ and $\lim_{T \rightarrow \infty} C(\infty) = 0$. It is a trivial matter to show that this is the only possible solution for $T \rightarrow \infty$.

In the case where $g[u] = \text{sgn}[u]$ we can push the analysis of (26,27) further for arbitrary T , as we may do the integrations in (26) explicitly (in the recall phase this would not have been possible). For $g[u] = \text{sgn}[u]$ we may write

$$C(\tau) = \int Dx Dy \text{sgn}[x + \tan(\psi)y] \text{sgn}[x - \tan(\psi)y] \quad (29)$$

$$\tan^2(\psi) = [\Xi(0) - \Xi(\tau)] / [\Xi(0) + \Xi(\tau)] \quad (30)$$

Upon writing (x, y) in polar coordinates, the integrations can be done, and we find

$$\sin\left(\frac{1}{2}\pi C(\tau)\right) = \frac{T e^{-|\tau|} + \frac{\alpha}{2} \int_{-\infty}^{\infty} du e^{-|u+\tau|} C(u)}{T + \frac{\alpha}{2} \int_{-\infty}^{\infty} du e^{-|u|} C(u)} \quad (31)$$

For $T \rightarrow \infty$ we recover (28). Putting $\tau \rightarrow \infty$ in (31) gives

$$\sin\left(\frac{1}{2}\pi C(\infty)\right) = \frac{\alpha C(\infty)}{T + \alpha \int_0^{\infty} du e^{-u} C(u)} \quad (32)$$

The paramagnetic (P) state $C(\infty) = 0$ always solves (31), as expected. It also follows from (31) that a continuous transition from a paramagnetic to a spin-glass (SG) state $\{m = 0, C(\infty) > 0\}$ would occur (unless preceded by a recall transition) at $\kappa_c = 2\alpha/\pi$.

5.2. Detailed balance type solutions at zero temperature

Let inspect whether and where we have detailed balance type macroscopic laws, where after elimination of transient parts one finds ‘effective’ equations with $C(\tau > 0) = q$. Here the ansatz $C(\tau > 0) = q$ implies that $\Xi(\tau) = Te^{-|\tau|} + \alpha q$ and $\kappa = T + \alpha q$. One now finds that it gives proper solutions of our order parameter equations if and only if $T = 0$, in which case we obtain (for $\theta = 0$) $q_0 = q$ and the familiar looking

$$m = \int Dx g[m + x\sqrt{\alpha q}] \quad q = \int Dx g^2[m + x\sqrt{\alpha q}] \quad (33)$$

For $g[u] = \tanh[\beta u]$ these equations (33) are identical to those of the *symmetrically* extremely diluted attractor network with binary neurons [9]. For $g[u] = \text{sgn}[u]$ we immediately find the three possible solutions P $\{m = 0, q = 0\}$, SG $\{m = 0, q = 1\}$ (both P and SG exist for any α) and R $\{m = m^*, q = 1\}$, where m^* is the solution of

$$m = \text{Erf}[m/\sqrt{2\alpha}] \quad (34)$$

The retrieval solution bifurcates from the SG one at $\alpha_c = 2/\pi \approx 0.637$.

The constant correlation function implies that the spins are microscopically frozen, thus there are no probability currents at all in the system so detailed balance is trivially restored. It is clear that such solutions always exist and are stable at $\alpha = 0$.

We next inspect the local stability of the above special $T = 0$ solutions against perturbations away from $C(\tau > 0) = q$ for $\alpha > 0$. We thus put $m \rightarrow m + \delta m$ and $\Xi(\tau) \rightarrow q + \delta\Xi(\tau)$ (with δm and $\delta\Xi(\tau)$ small) and find the eigenvalue problem

$$\delta\Xi(\tau) = K_1 \int_{-\infty}^{\infty} du \delta\Xi(u) \left[\frac{K_2}{1 - K_2} e^{-|u|} + e^{-|\tau - u|} \right] \quad (35)$$

where, with $y(x) = m + x\sqrt{\alpha q}$

$$K_1 = \frac{1}{2}\alpha \int Dx \left\{ g'[y] \right\}^2 \quad (36)$$

$$K_2 = \alpha \int Dx g[y]g''[y] + \frac{\sqrt{\frac{\alpha}{q}} \int Dx g'[y] \cdot \int Dx g[y]g''[y]}{1 - \int Dx g'[y]} \quad (37)$$

The only eigenfunction of (35) which is not trivial and bounded for $\tau \rightarrow \infty$ is given by

$$\delta\Xi(\tau) = \cos(\sqrt{2K_1 - 1}\tau) - \frac{K_2}{1 - K_2 - 2K_1} \quad (38)$$

with $K_1 \geq \frac{1}{2}$. Since $\lim_{\alpha \rightarrow 0} K_1 = 0$, the bifurcation condition becomes $K_1 = \frac{1}{2}$, or

$$1 = \alpha \int Dx \left\{ g'[m + x\sqrt{\alpha q}] \right\}^2 \quad (39)$$

The size of the region along the $T = 0$ axis where the solution (33) is stable is dependent on our choice of the gain function g . For $g[u] = \text{sgn}[u]$ we find that there is never a continuous transition while for $g[u] = \tanh[\gamma u]$ there can be and the solution (33) may be stable in a large part of the recall region. To see more clearly why there are no

continuous transitions for $g[u] = \text{sgn}[u]$ we note that in that case the special solution is $q = q_0 = 1$, and our eigenvalue problem becomes

$$\alpha + \delta\Xi(\tau) = \frac{\alpha}{2} \int_{-\infty}^{\infty} du e^{-|u+\tau|} \int Dx \text{Erf}^2 \left(\frac{m + x\sqrt{\alpha + \delta\Xi(u)}}{y\sqrt{\delta\Xi(0) - \delta\Xi(u)}} \right) \quad (40)$$

For small perturbations the argument of the error function will be large and we can use the expansion [23]

$$\text{Erf}(x) = 1 - \frac{1}{\sqrt{\pi}x} e^{-x^2} \sum_{n=0}^{\infty} \left(1 - \frac{1}{2x^2} + \frac{1.3}{(2x^2)^2} - \frac{1.3.5}{(2x^2)^3} + \dots \right) \quad (41)$$

to see that on the right-hand side of (40) there is no polynomial term in the perturbation, and hence no continuous bifurcation.

6. Non-equilibrium phase transitions

6.1. Continuous transitions without anomalous response

Continuous transitions from P or SG states to recall states (R, where $m > 0$) or from P states to SG states, are found upon expanding either (23) or (24) (both with $\theta = 0$) for small m or small q , respectively. This is found to give the simple bifurcation conditions

$$P \rightarrow R, \text{ SG} \rightarrow R : \quad \int Dx g'[x\sqrt{\kappa}] = 1 \quad (42)$$

$$P \rightarrow \text{SG} : \quad \int Dx g'[x\sqrt{\kappa}] = 1/\sqrt{\alpha} \quad (43)$$

For $g[u] = \text{sgn}[u]$ these conditions reduce to $\kappa_c = 2/\pi$ and $\kappa_c = 2\alpha/\pi$, respectively. Here, since $\kappa \in [T, T + \alpha q_0]$, we are sure that continuous transitions $P \rightarrow R$ or $\text{SG} \rightarrow R$ can occur at most for $T \in [2/\pi - \alpha, 2/\pi]$, and that continuous transitions $P \rightarrow \text{SG}$ can occur only when $T < 2\alpha/\pi$. To determine which instability away from the paramagnetic state occurs first we need to inspect the dependence of κ , as defined in (25) on (α, T) .

6.2. Transitions to states with anomalous response

We next inspect the solution of our full TTI equations for large times, upon separating non-persistent from persistent terms in the functions $\Xi(\tau)$ from $C(\tau)$:

$$C(\tau) = q + \tilde{C}(\tau), \quad \Xi(\tau) = \alpha q + \tilde{\Xi}(\tau) \quad (44)$$

Insertion into our equations for $\Xi(\tau)$ and $C(\tau)$ gives

$$\tilde{\Xi}(\tau) = T e^{-|\tau|} + \frac{1}{2}\alpha \int_{-\infty}^{\infty} du e^{-|u+\tau|} \tilde{C}(u) \quad (45)$$

$$\tilde{C}(\tau) = \int Dx \left\{ \int Dy g \left[m + x\sqrt{\alpha q + \tilde{\Xi}(\tau)} + y\sqrt{\kappa - \alpha q - \tilde{\Xi}(\tau)} \right] \right\}^2 - q \quad (46)$$

For large times the non-persistent terms will be small, so we may expand the right-hand side of (46) in powers of $\tilde{\Xi}(\tau)$. First we work out the nonlinearity $g[\dots]$:

$$g[\dots] = g[m + x\sqrt{\alpha q} + y\sqrt{\kappa - \alpha q}] + \frac{1}{2}\tilde{\Xi}(\tau) \left[\frac{x}{\sqrt{\alpha q}} - \frac{y}{\sqrt{\kappa - \alpha q}} \right] g'[m + x\sqrt{\alpha q} + y\sqrt{\kappa - \alpha q}] + \mathcal{O}(\tilde{\Xi}^2) \quad (47)$$

which gives, after partial integration over the Gaussian disorder variables,

$$\tilde{C}(\tau) = \Lambda \tilde{\Xi}(\tau) + \mathcal{O}(\tilde{\Xi}^2(\tau)) \quad (48)$$

$$\begin{aligned} \Lambda &= \int Dx \left\{ \int Dy g[m + x\sqrt{\alpha q} + y\sqrt{\kappa - \alpha q}] \right\} \\ &\quad \times \left\{ \int Dy \left[\frac{x}{\sqrt{\alpha q}} - \frac{y}{\sqrt{\kappa - \alpha q}} \right] g'[m + x\sqrt{\alpha q} + y\sqrt{\kappa - \alpha q}] \right\} \\ &= \int Dx \left\{ \int Dy g'[m + x\sqrt{\alpha q} + y\sqrt{\kappa - \alpha q}] \right\}^2 \end{aligned} \quad (49)$$

For sufficiently large times we may therefore replace the duo (45,46) by the leading order

$$\tilde{C}(\tau) = \Lambda \left\{ T e^{-|\tau|} + \frac{1}{2}\alpha \int_{-\infty}^{\infty} du e^{-|u+\tau|} \tilde{C}(u) \right\} + \dots \quad (50)$$

This equation is solved asymptotically by $\tilde{C}(\tau) \sim e^{-\gamma\tau}$, with exponent $\gamma = \sqrt{1 - \alpha\Lambda}$. We must conclude that a transition to a regime with anomalous response (which in equilibrium disordered systems would correspond to the AT line [21]) occurs when $\alpha\Lambda = 1$, i.e. when

$$\alpha \int Dx \left\{ \int Dy g'[m + x\sqrt{\alpha q} + y\sqrt{\kappa - \alpha q}] \right\}^2 = 1 \quad (51)$$

We see that in the paramagnetic state $m = q = 0$ the condition (51) coincides with that of the P→SG transition (43), and that for $T = 0$ it coincides with the instability condition (39). Again, for the special choice $g[u] = \text{sgn}[u]$ all integrals can be done analytically, and we find (51) taking the more explicit form

$$\alpha e^{-m^2/(\kappa + \alpha q)} = \frac{1}{2}\pi \sqrt{\kappa^2 - \alpha^2 q^2} \quad (52)$$

7. Bounds and approximations for the stationary state

7.1. Arbitrary non-linearities

We established that it is not possible to obtain closed equations for persistent order parameters only, due to the dependence of (25) on the short-time part of $C(\tau)$. However, we know that $C(\tau)$ decays from $C(0) = q_0$ to $C(\infty) = q$, so that (25) can be written as

$$\kappa = T + \alpha q + \alpha \Delta (q_0 - q), \quad \Delta \in [0, 1] \quad (53)$$

with $\Delta \approx 1$ when the decay of $C(\tau)$ is much slower than $\exp[-\tau]$, and Δ smaller when the decay is fast. Thus we know that our phase diagram must interpolate between those

obtained for the extreme cases $\Delta \in \{0, 1\}$, with these cases becoming identical both for $T \rightarrow 0$ at small α (where $q_0 - q \rightarrow 0$ before the transition to non-flat $C(\tau)$) and $\alpha \rightarrow 0$:

$$\text{'Slow' limit, } \Delta = 1 : \quad m = \int Dx \, g[m + x\sqrt{T + \alpha q_0}] \quad (54)$$

$$q_0 = \int Dx \, g^2[m + x\sqrt{T + \alpha q_0}] \quad (55)$$

$$q = \int Dx \left\{ \int Dy \, g[m + x\sqrt{\alpha q} + y\sqrt{T + \alpha(q_0 - q)}] \right\}^2 \quad (56)$$

$$\text{'Fast' limit, } \Delta = 0 : \quad m = \int Dx \, g[m + x\sqrt{T + \alpha q}] \quad (57)$$

$$q_0 = \int Dx \, g^2[m + x\sqrt{T + \alpha q}] \quad (58)$$

$$q = \int Dx \left\{ \int Dy \, g[m + x\sqrt{\alpha q} + y\sqrt{T}] \right\}^2 \quad (59)$$

Rather than analyze the extreme bounding cases $\Delta \in \{0, 1\}$, one could construct a rational interpolation between $\Delta = 0$ and $\Delta = 1$ by using our knowledge of the long-time behaviour of the non-persistent correlations as established in the derivation of (51), viz. $\tilde{C}(\tau) \sim e^{-\tau\sqrt{1-\alpha\Lambda}}$ with Λ as given in (49), to define the simplest function which satisfies both the long-time profile and the initial conditions $\tilde{C}(0) = q_0 - q$: $\tilde{C}(\tau) = (q_0 - q)e^{-\tau\sqrt{1-\alpha\Lambda}}$. This implies the following approximate expression for Δ :

$$\Delta = \int_0^\infty d\tau \, e^{-\tau[1+\sqrt{1-\alpha\Lambda}]} = \frac{1}{1 + \sqrt{1 - \alpha\Lambda}} \quad (60)$$

Equivalently:

$$\kappa = T + \alpha q + \frac{\alpha(q_0 - q)}{1 + \sqrt{1 - \alpha\Lambda}} \quad (61)$$

This then leads us to the following theory:

$$\text{interpolation : } m = \int Dx \, g\left[m + x\sqrt{T + \alpha q + \frac{\alpha(q_0 - q)}{1 + \sqrt{1 - \alpha\Lambda}}}\right] \quad (62)$$

$$q_0 = \int Dx \, g^2\left[m + x\sqrt{T + \alpha q + \frac{\alpha(q_0 - q)}{1 + \sqrt{1 - \alpha\Lambda}}}\right] \quad (63)$$

$$q = \int Dx \left\{ \int Dy \, g\left[m + x\sqrt{\alpha q} + y\sqrt{T + \frac{\alpha(q_0 - q)}{1 + \sqrt{1 - \alpha\Lambda}}}\right] \right\}^2 \quad (64)$$

$$\Lambda = \int Dx \left\{ \int Dy \, g'\left[m + x\sqrt{\alpha q} + y\sqrt{T + \frac{\alpha(q_0 - q)}{1 + \sqrt{1 - \alpha\Lambda}}}\right] \right\}^2 \quad (65)$$

Again we note that these equations are exact in either of the limits $T \rightarrow 0$ and $\alpha \rightarrow 0$.

7.2. Predictions for $g[u] = \text{sgn}[u]$

For the choice $g[u] = \text{sgn}[u]$, where Gaussian integrals can be done, we have $q_0 = 1$. Thus the slow limit gives the P \rightarrow F transition line $\kappa = \alpha + T = 2/\pi$. In the paramagnetic

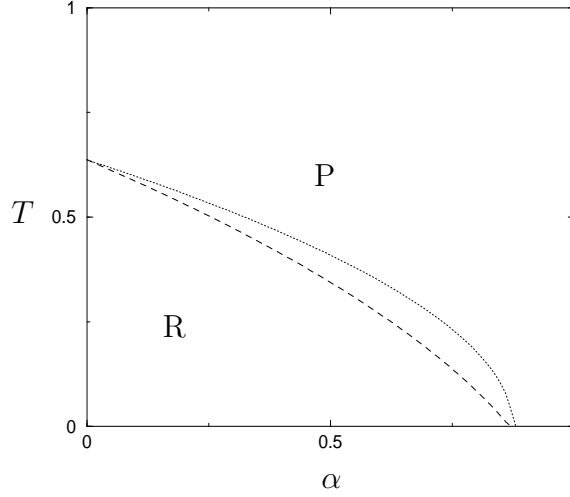


Figure 1. Phase diagram for the asymmetrically dilute graded response model. The dashed line is the transition predicted by the interpolation approximation while the dotted line is that predicted by numerically solving for the correlation function. The interpolative method gives reasonable agreement but systematically underestimates the size of the recall region with the error increasing as a function of α .

state $q = 0$ thus the fast limit gives a $P \rightarrow F$ transition for $\alpha < 1$ at $T = 2/\pi$ and a $P \rightarrow SG$ transition for $\alpha > 1$ at $T = 2\alpha/\pi$. In addition, the above set of interpolating equations can be simplified to

$$m = \text{Erf}[m/\sqrt{2\kappa}], \quad q = \int Dx \text{Erf}^2\left[\frac{m + x\sqrt{\alpha q}}{\sqrt{2(\kappa - \alpha q)}}\right] \quad (66)$$

$$\kappa = T + \alpha \left\{ \frac{(\kappa^2 - \alpha^2 q^2)^{\frac{1}{4}} + q(\sqrt{\kappa^2 - \alpha^2 q^2} - \frac{2\alpha}{\pi} e^{-m^2/(\kappa + \alpha q)})^{\frac{1}{2}}}{(\kappa^2 - \alpha^2 q^2)^{\frac{1}{4}} + (\sqrt{\kappa^2 - \alpha^2 q^2} - \frac{2\alpha}{\pi} e^{-m^2/(\kappa + \alpha q)})^{\frac{1}{2}}} \right\} \quad (67)$$

with the phase transitions occurring at:

$$P \rightarrow R, \text{ SG} \rightarrow R : \quad \kappa_c = \frac{2}{\pi} \quad (68)$$

$$P \rightarrow \text{SG} : \quad \kappa_c = \frac{2\alpha}{\pi} \quad (69)$$

$$\text{Anomalous Response} : \quad \sqrt{\kappa^2 - \alpha^2 q^2} = \frac{2\alpha}{\pi} e^{-m^2/(\kappa + \alpha q)} \quad (70)$$

In the paramagnetic state $m = q = 0$, where the conditions (69,70) become identical, equation (67) is seen to simplify significantly, and can in fact be solved for κ :

$$\kappa = \frac{T(1 - \frac{2}{\pi}) + \frac{1}{2}\alpha + \sqrt{T^2 + \alpha T(1 - \frac{2}{\pi}) + \frac{1}{4}\alpha^2}}{2(1 - \frac{1}{\pi})} \quad (71)$$

Given this expression we may in turn derive explicit expressions for the our phase transition transition lines. One finds that the transition $P \rightarrow \text{SG}$ cannot occur, and that the $P \rightarrow R$ transition line reduces to

$$P \rightarrow R : \quad T_c(\alpha) = \sqrt{1 - \alpha} - 1 + \frac{2}{\pi} \quad (72)$$

This critical temperature decreases with increasing α from its maximum value $T_c(0) = 2/\pi \approx 0.637$ down to $T_c(\alpha_c) = 0$, with $\alpha_c = \frac{4}{\pi}(1 - \frac{1}{\pi}) \approx 0.868$. Since our interpolation equations are exact at $T = 0$ and at $\alpha = 0$, we can be sure that also the two critical values $T_c(0) = 2/\pi$ and α_c are exact.

Thus, given that one always enters a recall phase directly (without a spin-glass phase, unlike symmetrically connected models), the only remaining transition to be investigated is the one marking the possible onset of anomalous response inside the retrieval phase. Combining (67) with (70) shows that at this transition $\kappa = T + \alpha$, so that we have to solve the transition line from the following trio of coupled equations

$$\text{Anomalous Response : } m = \text{Erf}\left[\frac{m}{\sqrt{2(T + \alpha)}}\right] \quad (73)$$

$$q = \int Dx \text{ Erf}^2\left[\frac{m + x\sqrt{\alpha q}}{\sqrt{2[T + \alpha(1 - q)]}}\right] \quad (74)$$

$$[T + \alpha(1 - q)][T + \alpha(1 + q)] = \frac{4\alpha^2}{\pi^2} e^{-\frac{2m^2}{T + \alpha(1 + q)}} \quad (75)$$

Examining (25) we see that for an anomalous response we require $q = 1$ which via (74,75) implies $T = \alpha = 0$ is the only possibility. However, there one trivially has $C(\tau) = 1, \forall \tau$ and there is no transient term in the correlation function.

We plot the phase diagram Figure 1, both using the approximation (72) and by solving for the whole correlation function numerically. We see that the results are similar although not identical, the interpolation method seems to predicts a slightly smaller recall region.

8. Comparison with simulations

In Figures 2 and 3 we examine both the distribution of spins at a single time and the correlation function. In simulations the spins evolved for $\mathcal{O}(10^5)$ discretized time steps (with $\Delta t = 0.02$) in order that they reached their static distribution. It is clear from (13) that the static distribution of spins must be Gaussian with mean m and variance κ . Our two theories (exact versus interpolation) give different values for m and κ , yet both are seen to agree very well with the simulations as far as the spin distribution is concerned. For the correlation function, however, it appears that solving the full order parameter equations numerically gives excellent agreement with simulations and is a significantly better guide to both the persistent correlation and the short time shape than interpolation. The differences between the two theories appear to increase with α within the recall regime.

In Figure 4 we plot magnetisations m and persistent correlations q for a variety of values for T and α on both sides of the predicted phase boundary, for both simulations and theory. In the left picture we see that for moderate α ($= 0.2$) both theories give good agreement with each other and with simulations. However, in the right picture, i.e. at higher α near the phase transition, again the numerical solution of the full order parameter equations gives better agreement.

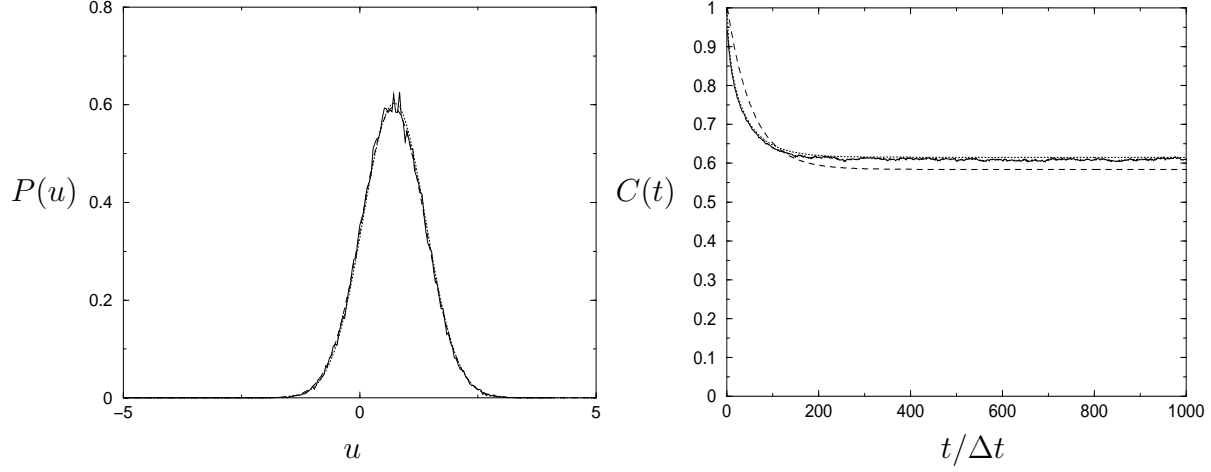


Figure 2. The left graph depicts the static distribution of spins at $\alpha = T = 0.25$. The solid line is from a simulation of $N = 64000$ spins, the dotted line is from solving for the correlation function and the dashed line is from the interpolation theory. Both theories give excellent agreement with the simulation. The right graph depicts the corresponding TTI correlation function as a function of the number of discretized time steps $t/\Delta t$ (with here $\Delta t = 0.02$), with the meaning of the different lines are as in the left picture. It is clear that the solving for the correlation function numerically gives a significantly better calculation of the correlation function.

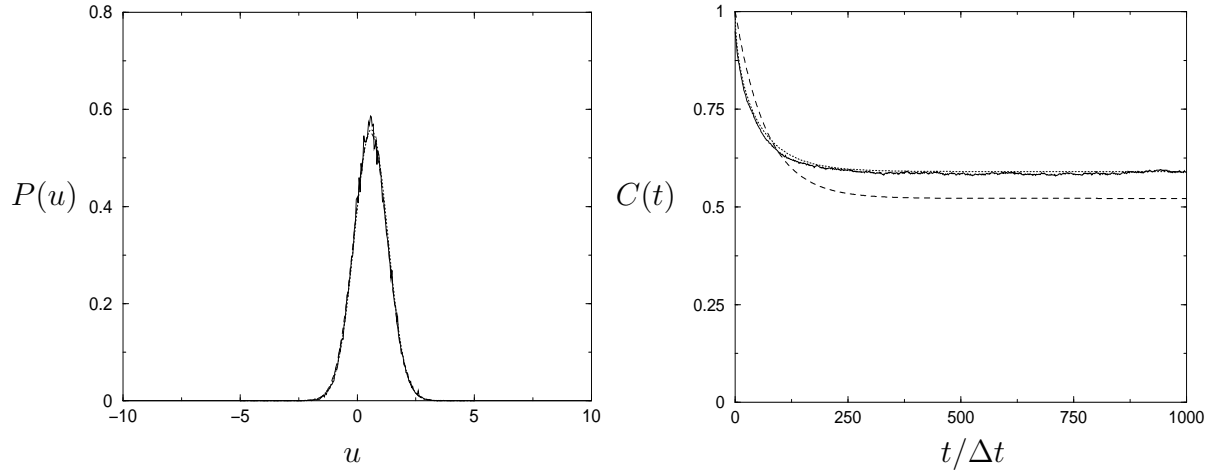


Figure 3. The left graph depicts the static distribution of spins at $\alpha = 0.5$ and $T = 0.125$. The solid line is from a simulation of $N = 64000$ spins, the dotted line is from solving for the correlation function and the dashed line is from the interpolation theory. Both theories give excellent agreement with the simulation. The right graph depicts the corresponding TTI correlation function as a function of the number of discretized time steps $t/\Delta t$ (with here $\Delta t = 0.02$), with the meaning of the different lines are as in the left picture. It is clear that the solving for the correlation function numerically gives a significantly better calculation of the correlation function.

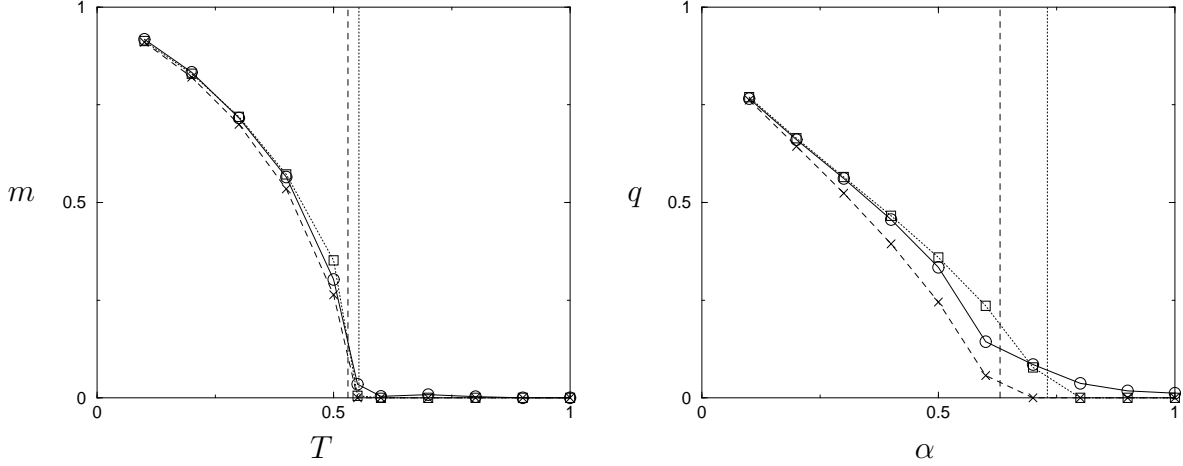


Figure 4. The left graph shows the magnetisation versus temperature for $\alpha = 0.2$. The solid line with \circ markers is averaged over 5 simulations of $N = 64000$ spins (error bars are all < 0.01), the dotted line with \square markers is obtained via solving the correlation function numerically, and the dashed line with \times markers is calculated from the interpolation theory. The vertical lines mark the predicted phase transitions, the left line being from the interpolation. Both theories give excellent agreement with the simulation. The right graph shows the persistent correlation versus α at $T = 0.25$. The different lines are as above. Solving the correlation function numerically gives better agreement with the simulations at higher α .

9. Discussion

In this paper we have used the formalism of generating functional theory to solve the dynamics of an extremely dilute asymmetrically structured Hopfield-type attractor neural network model with Langevin dynamics near saturation. This is a non-detailed-balance problem. In contrast to extremely diluted models with synchronous dynamics, we here cannot benefit from the Gaussian fields since we need more than just the marginal field distribution. We have looked for time translationally invariant solutions as an aid to finding stationary solutions to this problem. In general we find that we need a full functional order parameter $C(\tau)$ to describe this stationary solution exactly. We have also looked at an interpolative theory that, while only exact at $\alpha = 0$, depends on only a few scalar parameters rather than an entire order parameter function. This reduced theory gives good agreement at low α but shows discrepancies at higher values of α .

As expected for this type of architecture the phase diagrams exhibit just two phases, a recall phase and a non-recall phase. There are no spin glass phases, which is due to the lack of a retarded self-interaction in the effective single spin equation. We have calculated, for specific choices of the gain function g , the system's (α, T) phase diagram, and have found the second order transitions in this diagram via a bifurcation analysis. The methods used in this paper to solve the effective single spin equation could be valid for other problems with a similar architecture and similar dynamics. The

network's connection asymmetry is obviously an assumption whose justification is model dependent, here it is reasonable and leads to exactly solvable statics.

It would also be of interest to study the initial dynamics of this problem, as that will be of more relevance to biological experiments. Other generalizations could include moving to a finitely connected architecture which may be closer to that seen in nature. Here the order parameters would become single spin path probabilities, and the effect on these path probabilities of a change in the external field at each point on the path [24]. It may also be of interest to see if an approximate theory could be constructed when arbitrary degrees of symmetry are introduced into the problem, such that the method presently used in this paper is no longer accessible.

References

- [1] Hopfield JJ 1982 *Proc. Natl. Acad. Sci. USA* **79** 2554
- [2] Hopfield JJ 1984 *Proc. Natl. Acad. Sci. USA* **81** 3088
- [3] Amit DJ, Gutfreund H and Sompolinsky H 1985 *Phys. Rev. A* **32** 1007
- [4] Amit DJ, Gutfreund H and Sompolinsky H 1985 *Phys. Rev. Lett.* **55**
- [5] Rieger H, Schreckenberg M and Zittartz J 1988 *Z. Phys. B* **72** 523
- [6] Horner H, Bormann D, Frick M, Kinzelbach H and Schmidt A 1989 *Z. Phys. B* **76** 383
- [7] Derrida B, Gardner E, Zippelius A, 1987 *Europhys. Lett.*, Vol 4 (2) pg 167-173
- [8] Kree R and Zippelius A 1991 in *Models of Neural Networks I* Domany R, van Hemmen JL and Schulten K (Eds) (Berlin: Springer) 193
- [9] Watkin TLH and Sherrington D 1991 *J. Phys. A: Math. Gen.* **24** 5427
- [10] Mertens S 1991 *J. Phys. A: Math. Gen.* **24** 337
- [11] B Wemmenhove B and Coolen A C C 2003 *J. Phys. A: Math. Gen.* **36** 9617
- [12] Kühn R 1991 *Statistical Mechanics of Neural Networks (Springer Lecture Notes in Physics 398)* (Heidelberg: Springer; ed L Garrido) p19
- [13] Kühn R, Bös S and Van Hemmen JL 1991 *Phys. Rev. A* **43** 2084
- [14] Kühn R and Bös S 1993 *J. Phys. A: Math. Gen.* **26** 831
- [15] Coolen ACC 2001 in *Handbook of Biological Physics, Vol 4* (Elsevier Science; eds. F.Moss and S. Gielen) 597
- [16] De Dominicis C 1978 *Phys. Rev. B* **18** 4913
- [17] Crisanti A and Sompolinsky H 1987 *Phys. Rev. A* **36** 4922
- [18] Crisanti A and Sompolinsky H 1988 *Phys. Rev. A* **37** 4865
- [19] Rieger H, Schreckenberg M and Zittartz J 1989 *Z. Phys. B* **74** 527
- [20] Düring A, Coolen ACC and Sherrington D 1998 *J. Phys. A: Math. Gen.* **31** 8607
- [21] de Almeida J R L and Thouless D J 1978 *J. Phys. A: Math. Gen.* **11** 983
- [22] Eissfeller H and Oppen M 1994 *Phys. Rev. E* **50** 709
- [23] Menzel D (Ed.) 1960 *Fundamental Formulas of Physics* (New York: Dover)
- [24] Hatchett J P L, Wemmenhove B, Pérez-Castillo I, Nikolettopoulos T, Skantzos N S and Coolen A C C 2004, preprint `cond-mat/0403172`